Tuning indirect predictions based on SNP effects from ssGBLUP

Daniela Lourenco
A. Legarra, S. Tsuruta, D. Moser, S. Miller, I. Misztal

Interbull 2018
Why Indirect predictions?

• Interim evaluations
  • Between official runs

• Not all genotyped animals are in the evaluations
  • Animals with incomplete pedigree increase bias and lower $R^2$

• Commercial products
  • e.g. GeneMax for non-registered animals
Indirect predictions in ssGBLUP

\[ \begin{bmatrix} X'X & X'W \\ W'X & W'W + H^{-1}\lambda \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{u} \end{bmatrix} = \begin{bmatrix} X'y \\ W'y \end{bmatrix} \]

\[ DGV = Z\hat{a} \]

\[ \hat{a} = \lambda D Z'G^{-1} \hat{u} \]

\[ G_{\text{APY}}^{-1} = A_{22}^{-1} + G^{-1} \]

\[ G_{\text{EGBVs}}^{-1} = \frac{1}{\lambda} D G V = \frac{1}{\lambda} Z a \]

\[ \text{GEBV}_{\text{young}} = w_1 PA + w_2 DGV - w_3 PP \]

\[ \text{GEBV}_{\text{young}} \approx \text{DGV} = Z\hat{a} \]

Lourenco et al., 2015
Problems with Indirect predictions

\[ \text{COR(GEBV, } \hat{Z}_a) > 0.99 \]

\[ \text{Avg(GEBV)} \approx 100 \quad \Rightarrow \quad \text{Avg}(\hat{Z}_a) \approx 0 \]
Objectives

1) Fine-tune indirect predictions to be compatible with GEBV

2) Investigate whether SNP effects are accurate when APY is used
   • Possibly use subset of core animals
Dataset

- American Angus Association
  - 8.2M animals in pedigree
  - 6.2M birth weight (BW)
  - 6.8M weaning weight (WW)
  - 3.4M post-weaning gain (PWG)
  - 81k genotyped
    - born 1977-2012: 66k
    - born 2013-2014: 15k
- Complete
  - Phenotypes up to 2012
  - Genotypes up to 2014 (81k)
- Reduced
  - Phenotypes up to 2012
  - Genotypes up to 2012 (66k)
- 3-trait with mat and mpe
  - Results for PWG
Accuracy of SNP effects from $G_{APY}^{-1}$ or $G_{cc}^{-1}$

$\hat{a}_G = \lambda D Z' G^{-1} \hat{u}$

$\hat{a}_{G_{APY}^{-1} H} = \lambda D Z' G_{APY\_high\_reliability}^{-1} \hat{u}_{APY}$

$\hat{a}_{G_{APY}^{-1} R} = \lambda D Z' G_{APY\_random}^{-1} \hat{u}_{APY}$

$\hat{a}_{G_{cc}^{-1} H} = \lambda D Z' G_{cc\_high\_reliability}^{-1} \hat{u}_{APY}$

$\hat{a}_{G_{cc}^{-1} R} = \lambda D Z' G_{cc\_random}^{-1} \hat{u}_{APY}$

- Correlation between SNP effects
- Correlation between $Z\hat{a}$
Statistics for SNP effects

\[ G^{-1} \quad >0.99 \quad >0.99 \]

\[ G_{\text{APY}_{\text{High}}}^{-1} \quad >0.99 \]

\[ G_{\text{APY}_{\text{Rand}}}^{-1} \]
Statistics for SNP effects

\[
\begin{align*}
G^{-1} & \quad 0.93 & \quad 0.90 \\
G_{CC\_High}^{-1} & \quad 0.90 \\
G_{CC\_Rand}^{-1} & \\
\end{align*}
\]
Statistics for $Z\hat{a}$

$G^{-1}$ 0.989 0.988

$G_{CC\_High}^{-1}$ 0.987

$G_{CC\_Rand}^{-1}$
Fine-tuning indirect predictions from ssGBLUP

Understanding genetic and genomic bases

• Base of BLUP: *founders of the pedigree*

• Base of GBLUP: *genotyped* animals

• Base of SSGBLUP: Vitezica et al. (2011) modeled as a mean in genotyped animals

  • \( p(u_g) = N(1\mu, G) \)

  • \( \mu = \text{(Pedigree base)} - \text{(Genomic base)} \)
Fine-tuning indirect predictions from ssGBLUP

1) Formula in Legarra (2017)
\[ \hat{u}_{ip} = \hat{\mu} + 0.95\hat{Z}\hat{a} + 0.05 \hat{u}_{parents} \]

2) Double fitting
   a) fit a regression using genotyped animals in the evaluation
   \[ \text{DGV}_{eval} = b_0 + b_1 Z\hat{a} \]
   b) apply regression for indirectly predicted animals
   \[ \hat{u}_{ip} = b_0 + b_1 Z\hat{a} \]

3) Add average GEBV
   \[ \hat{u}_{ip} = \overline{\text{GEBV}}_{eval} + Z\hat{a} \]
Bias of indirect predictions

GEBV – $u_{ip}$
Correlation & Regression Coefficient

Za
Legarra_2017
Double_Fit
Average_GEBV

Correlation(GEBV,u_{ip})

Regression of GEBV on u_{ip}

Fine-tuning indirect predictions in ssGBLUP

\[
E(\hat{u}|\hat{a}) = \mu + Z\frac{1}{2\sum p(1-p)}\left(I\frac{1}{2\sum p(1-p)}\right)^{-1}(\hat{a} - 0)
\]

\[
E(\hat{u}|\hat{a}) = \mu + Z\hat{a}
\approx
\]

\[
E(\hat{u}|\hat{a}) = \overline{GEBV} + Z\hat{a}
\]
Final Remarks

• Indirect predictions are unbiased after corrections
  • Can be used as interim evaluation

• Indirect predictions based on core animals are slightly less accurate
  • Reduction in computing time (no $G_{nc}^{-1}$ and $G_{nn}^{-1}$)

• SNP effects from ssGBLUP may be useful for SNP MACE
Acknowledgements